

AD-A112 920 CLEMSON UNIV SC DEPT OF MATHEMATICAL SCIENCES F/G 12/1
AN EMPIRICAL BAYES ESTIMATE OF MULTINOMIAL PROBABILITIES.(U)
FEB 82 K ALAM, A MITRA N00014-75-C-0451

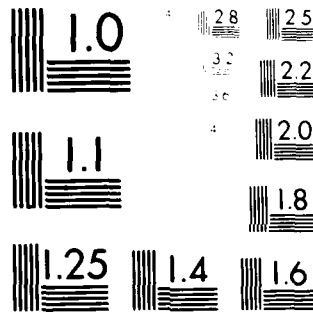
UNCLASSIFIED N137

NL

1 111 1
A112
11111



END
DATE
FILMED
104-82
DTIC



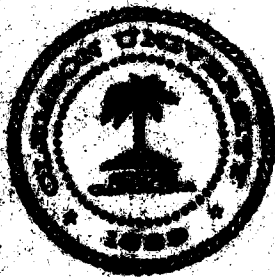
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

12

AD 1113-000

DEPARTMENT
OF
MATHEMATICAL
SCIENCES

CLEMSON UNIVERSITY
Clemson, South Carolina



FREE COPY

DTIC
SELECTED
PT 1

REPRODUCTION PROHIBITED
Approved for public release
Distribution Unlimited

82 04 05 028

12

AN EMPIRICAL BAYES ESTIMATE
OF MULTINOMIAL PROBABILITIES

*Khursheed Alam
Clemson University

&

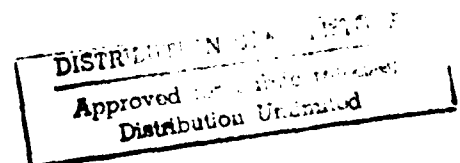
Amitava Mitra
Auburn University

N-137

Technical Report #381
Department of Mathematical Sciences
Clemson University

February, 1982

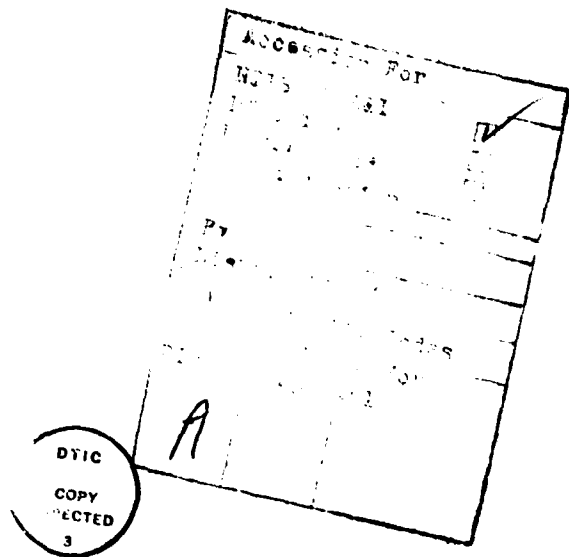
*This work was supported by the U. S. Office of Naval
Research under Contract N00014-75-C-0451.



An Empirical Bayes Estimate of Multinomial Probabilities

Khursheed Alam* & Amitava Mitra

Clemson University & Auburn University



*This work was supported by the U.S. Office of Naval Research under contract N00014-75-C-0451.

An Empirical Bayes Estimate of Multinomial Probabilities

Khursheed Alam & Amitava Mitra

Clemson University & Auburn University

ABSTRACT

This paper deals with the problem of estimating simultaneously the parameters (cell probabilities) of $m \geq 2$ independent multinomial distributions, with respect to a quadratic loss function. An empirical Bayes estimator is given which is shown to have smaller risk than the maximum likelihood estimator for sufficiently large values of m , at all points of the parameter space outside its boundary.

Key words: Multinomial distribution; Maximum likelihood; Empirical Bayes;
Quadratic loss

AMS Classification: 62F10

This work was supported by the U.S. Office of Naval Research under contract N00014-75-C-0451.

1. INTRODUCTION

The multinomial distribution arises in various statistical analyses. In this paper we consider the problem of estimating simultaneously the parameters (cell probabilities) of $m \geq 2$ independent multinomial distributions. A new estimator is obtained from empirical Bayes considerations, where the estimate of the parameters of one distribution depends on the observations taken from the other distributions. We compare the new estimator with the maximum likelihood estimator (MLE) and show that for large values of m the risk of the new estimator is smaller than the risk of the MLE at all points excluding a small subset of the parameter space.

The problem of estimating the parameters of a multinomial distribution has been considered by various authors. Alam (1979) and Johnson (1971) have shown that for $m = 1$ the MLE is admissible with respect to the sum of squared errors as the loss function. The estimation of multinomial probabilities has been considered from a decision theoretic point of view by Steinhaus (1957), Trybula (1958) and Rutkowska (1977). In a recent paper Olkin and Sobel (1979) have discussed the problem of estimating the parameters of $m \geq 2$ binomial distributions.

In Section 2 we introduce the new empirical Bayes estimator. In Section 3 we compare the risk of the empirical Bayes estimator with the MLE.

2. EMPIRICAL BAYES ESTIMATOR

Let $\mathbf{x}_i = (x_{i1}, \dots, x_{iK})$, $i = 1, \dots, m$ denote independent observations from m multinomial distributions, each with $K \geq 2$ cells, where $\sum_{j=1}^K x_{ij} = n$. Let $\mathbf{p}_i = (p_{i1}, \dots, p_{iK})$ denote the vector of cell probabilities associated with the i th distribution, where $p_{ij} > 0$, $j = 1, \dots, K$ and $\sum_{j=1}^K p_{ij} = 1$.

For estimating the cell probabilities simultaneously, let the loss function be given by

$$L(\delta, P) = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^K (\delta_{ij} - p_{ij})^2$$

where δ_{ij} denotes the estimate of p_{ij} , $\delta = (\delta_{ij})$ and $P = (p_{ij})$.

First we obtain a Bayes estimator of P , as follows. Suppose that the parameters of the m distributions are distributed a priori, independently and identically according to the Dirichlet distribution, given by the density function

$$f(\underline{\theta}, \lambda) = \frac{\Gamma(K\lambda)}{(\Gamma(\lambda))^K} (\theta_1 \dots \theta_K)^{\lambda-1}$$

where λ is a positive number, $0 < \theta_j < 1$, $j = 1, \dots, K$ and $\sum_{j=1}^K \theta_j = 1$.

A Bayes estimator of P with respect to the Dirichlet prior distribution is given by $\delta^* = (\delta_{ij}^*)$, where

$$\delta_{ij}^* = (x_{ij} + \lambda) / (n + K\lambda). \quad (2.1)$$

Let $q_i = 1 - \sum_{j=1}^K p_{ij}^2$, $q = \frac{1}{m} \sum_{i=1}^m q_i$ and $y = \sum_{i=1}^m \sum_{j=1}^K x_{ij}^2$. The risk of δ^* is given by

$$R(\delta^*) = [(n - \lambda^2 K^2)q + K(K-1)\lambda^2] (n + K\lambda)^{-2}. \quad (2.2)$$

The value of λ minimizing $R(\delta^*)$ is equal to λ_0 , given by

$$\lambda_0 = q(K-1-Kq)^{-1}. \quad (2.3)$$

Let R^* denote the value of $R(\delta^*)$ for $\lambda = \lambda_0$. We have

$$R^* = \frac{q}{n + K\lambda_0} < \frac{q}{n} = R(\delta^0) \quad (2.4)$$

where δ^0 denotes the MLE, given by $\delta_{ij}^0 = x_{ij}/n$. An unbiased estimator of q is given by $\hat{q} = (mn^2 - y)/mn(n-1)$. Substituting \hat{q} for q in (2.3) we get an estimated value of λ , given by

$$\begin{aligned}\hat{\lambda} &= \hat{q}(K-1-K\hat{q})^{-1} \\ &= \infty \text{ for } \hat{q} = (K-1)/K.\end{aligned}\tag{2.5}$$

Substituting $\hat{\lambda}$ for λ in (2.1), we obtain an empirical Bayes estimator $\hat{\delta}$ of P , given by

$$\begin{aligned}\hat{\delta}_{ij} &= (x_{ij} + \hat{\lambda}) / (n + K\hat{\lambda}) \\ &= 1/K \text{ for } \hat{\lambda} = \infty.\end{aligned}\tag{2.6}$$

The quantity q_i is a measure of diversity of the i th multinomial distribution. It is equal to zero when the probability mass of the distribution is concentrated in a single cell. It achieves its maximum value equal to $(K-1)/K$ when the probability mass is equally distributed among the K cells of the distribution. The value of q represents the average diversity of the m multinomial distributions. In the following section we show that the empirical Bayes estimator $\hat{\delta}$ has smaller risk than the MLE for sufficiently large values of mq^2 . Therefore, in practice when the parameters of a large number of multinomial distributions are simultaneously estimated for which the average diversity is unlikely to be very small, the empirical Bayes estimator should be preferred to the MLE.

The preceding observation is supported by the numerical results shown in Table 1 below. The table gives the empirical values of $R(\hat{\delta})$, $R(\delta^0)$ and mq^2 for certain values of m , n , K and randomly chosen values of P . The value of $R(\hat{\delta})$ and $R(\delta^0)$ for any fixed value of P was obtained from simulation by the Monte Carlo method. It is seen from the table that $R(\hat{\delta}) < R(\delta^0)$ in all

the observed cases except one when $m = 5$, $n = 10$, $K = 2$ for which $R(\hat{\delta}) = .0247$ and $R(\delta^0) = .0241$. This case corresponds to the smallest value of $mq^2 = 0.2856$ given in the table.

3. RISK OF $\hat{\delta}$

From (2.2) it is seen that $R(\delta^*)$ first decreases then increases as λ varies from 0 to ∞ . It is equal to $\frac{q}{n}$ and $\frac{K-1}{K} - q$ for $\lambda = 0$ and ∞ , respectively. Hence, $R(\delta^*) < R(\delta^0)$ for all positive values of λ if $\frac{q}{n} > \frac{K-1}{K} - q$. Therefore, we shall assume that

$$q \leq \frac{n(K-1)}{K(n+1)}. \quad (3.1)$$

The value of $R(\hat{\delta})$ for large m is given as follows: From (2.3) and (2.5) we get

$$\hat{\lambda} - \lambda_0 = \frac{(K-1)(\hat{q}-q)}{(K-1-K\hat{q})(K-1-Kq)}. \quad (3.2)$$

Now, $E\hat{q} = q$ and $\text{Var}(\hat{q}) \leq \frac{nq}{m(n-1)}$ by Lemma 3.1 below. Therefore, for large m

$$E\hat{\lambda} = \lambda_0 + [K(K-1)/(K-1-Kq)^2] \text{Var}(\hat{q}) + o(\frac{1}{m}) \quad \text{and} \quad (3.3)$$

$$E(\hat{\lambda} - \lambda_0)^2 = [(K-1)^2/(K-1-Kq)^4] \text{Var}(\hat{q}) + o(\frac{1}{m}). \quad (3.4)$$

The value of $R(\hat{\delta})$ as obtained from (2.6) by direct computation, is given after simplification by

$$\begin{aligned} R(\hat{\delta}) = & R^* + nK(n+K\lambda_0)^{-2} E(n+K\hat{\lambda})^{-2} (n(K-1)-K(n-1)\hat{q})(\hat{\lambda}-\lambda_0)^2 \quad (3.5) \\ & + 2nK(n+K\lambda_0)^{-1} E(n+K\hat{\lambda})^{-1} (\hat{\lambda}-\lambda_0) \left\{ \frac{(K-1)\lambda_0 + (n-1)\hat{q} - (n+K\lambda_0)q}{n+K\lambda_0} + Z \right\} \end{aligned}$$

where $Z = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^K p_{ij}(x_{ij} - np_{ij})$. Now $EZ = 0$ and

$$\begin{aligned}
\text{Var}(Z) &= nm^{-2} \left(\sum_{i=1}^m \sum_{j=1}^K p_{ij}^3 - \sum_{i=1}^m (1-q_i)^2 \right) \\
&\leq nm^{-2} \sum_{i=1}^m q_i (1-q_i) \\
&\leq \frac{n}{m} q.
\end{aligned} \tag{3.6}$$

Using (3.3), (3.4), and (3.6) in (3.5) we get

$$\begin{aligned}
R(\hat{\delta}) &= R^* + \frac{cq}{m} + o\left(\frac{1}{m}\right) \\
&= q\left(\frac{1}{n+K\lambda_0} + \frac{c}{m}\right) + o\left(\frac{1}{m}\right)
\end{aligned} \tag{3.7}$$

where $c = c(P)$ is a number less than $K^2 n^2$ in absolute value. Then

$$\begin{aligned}
R(\delta^0) - R(\hat{\delta}) &= q\left(\frac{1}{n} - \frac{1}{n+K\lambda_0} - \frac{c}{m}\right) + o\left(\frac{1}{m}\right) \\
&= q\left(\frac{K\lambda_0}{n(n+K\lambda_0)} - \frac{c}{m}\right) + o\left(\frac{1}{m}\right) \\
&= q^2 \left[\frac{K}{n(n+K\lambda_0)}^{-1} (K-1-Kq)^{-1} - \frac{c}{mq} \right] + o\left(\frac{1}{m}\right).
\end{aligned} \tag{3.8}$$

By virtue of the inequality (3.1) the quantity inside the square bracket is minorized by

$$\frac{K}{2n^2(K-1)} - \frac{c}{mq}.$$

Therefore, we have the following result.

Theorem 3.1. The empirical Bayes estimator $\hat{\delta}$ has smaller risk than the maximum likelihood estimator δ^0 for sufficiently large values of mq^2 .

Remark. It has been mentioned in the introduction that the MLE is admissible with respect to the quadratic loss function when $m = 1$. It is also noted by Johnson (1971) that the admissibility of the MLE is essentially for the reason that its risk is small near the boundary of the parameter

space, given by $q = 0$. Therefore, the condition of Theorem 3.1 that the value of q should be sufficiently large for the risk of $\hat{\delta}$ to be smaller than the risk of δ^0 is consistent with that observation. Johnson has also noted that there is no Stein-effect for $m > 1$, that is, the MLE cannot be improved upon by letting the estimate of parameters of one distribution depend on the observations taken from the other distributions. Therefore, we do not expect that $\hat{\delta}$ has a uniformly smaller risk than δ^0 .

Finally, we give the lemma which was used in the proof of Theorem 3.1.

Lemma 3.1. $\text{Var}(\hat{q}) \leq \frac{nq}{m(n-1)}$

Proof: We have

$$\begin{aligned} \text{Var}\left(\sum_{j=1}^K x_{ij}^2\right) &= E\left(\sum_{j=1}^K x_{ij}^2\right)^2 - \left(E\sum_{j=1}^K x_{ij}^2\right)^2 \\ &\leq \left(E\sum_{j=1}^K x_{ij}^2\right)(n^2 - E\sum_{j=1}^K x_{ij}^2) \\ &\leq n^2(n^2 - E\sum_{j=1}^K x_{ij}^2) \\ &= n^3(n-1)q_i. \end{aligned}$$

Summing the above inequality for $i = 1, \dots, m$ we have

$$\text{Var}(y) \leq mn^3(n-1)q.$$

Therefore

$$\text{Var}(\hat{q}) \leq \frac{nq}{m(n-1)}. \quad (3.9)$$

References

- Alam, K. (1979), "Estimation of Multinomial Probabilities," Annals of Statistics, 7, 282-283.
- Johnson, B. McK. (1971), "On Admissible Estimators for Certain Fixed Sample Binomial Problems," Annals of Mathematical Statistics, 42, 1579-1587.
- Olkin, I., and Sobel, M. (1979), "Admissible and Minimax Estimation for the Multinomial Distribution and for Independent Binomial Distributions," Annals of Statistics, 7, 284-290.
- Rutkowska, M. (1977), "Minimax Estimation of the Parameters of the Multivariate Hypergeometric and Multinomial Distributions," Zastosowania Matematyki, 16, 9-21.
- Steinhaus, H. (1957), "The Problem of Estimation," Annals of Mathematical Statistics, 28, 633-648.
- Trybula, S. (1958), "Some Problems of Simultaneous Minimax Estimation," Annals of Mathematical Statistics, 245-253.

Table 1. Values of $R(\hat{\delta})$, $R(\delta^0)$ and mq^2

		5			10			15			20		
n	m	$R(\hat{\delta})$	$R(\delta^0)$	mq^2	$R(\hat{\delta})$	$R(\delta^0)$	mq^2	$R(\hat{\delta})$	$R(\delta^0)$	mq^2	$R(\hat{\delta})$	$R(\delta^0)$	mq^2
K = 2	5	0.0661	0.0669	0.5599	0.0582	0.0774	1.4394	0.0522	0.0668	1.6663	0.0516	0.0694	2.3633
	10	0.0247	0.0241	0.2856	0.0293	0.0311	0.9672	0.0301	0.0364	2.0446	0.0299	0.0362	2.5676
	15	0.0208	0.0217	0.5381	0.0229	0.0264	1.4815	0.0175	0.0187	1.1941	0.0211	0.0243	2.6616
	5	0.0797	0.0883	0.8782	0.0694	0.0951	2.3917	0.0661	0.1035	4.0701	0.0661	0.0875	3.9498
	10	0.0323	0.0338	0.5828	0.0339	0.0364	1.3454	0.0357	0.0408	2.5363	0.0386	0.0453	3.9980
	15	0.0272	0.0289	0.9254	0.0237	0.0254	1.5304	0.0261	0.0286	2.7626	0.0272	0.0307	4.1587
	5	0.0849	0.1111	1.5426	0.0773	0.1101	3.0063	0.0755	0.1054	4.2048	0.0739	0.1028	5.3127
	10	0.0480	0.0572	1.6411	0.0429	0.0482	2.2006	0.0672	0.0783	3.5883	0.0402	0.0461	4.2892
	15	0.0325	0.0369	1.5235	0.0304	0.0335	2.5538	0.0316	0.0368	4.5557	0.0293	0.0327	4.7678

Table 1. Values of $R(\hat{\delta})$, $R(\delta^0)$ and mq^2 (continued)

		30			40			50		
		$R(\hat{\delta})$	$R(\delta^0)$	mq^2	$R(\hat{\delta})$	$R(\delta^0)$	mq^2	$R(\hat{\delta})$	$R(\delta^0)$	mq^2
K = 2										
5	5	0.0502	0.0757	4.3331	0.0492	0.0667	4.6377	0.0494	0.0703	6.1646
10		0.0293	0.0346	3.5217	0.0289	0.0347	4.9293	0.0280	0.0321	4.9825
15		0.0180	0.0195	2.6324	0.0190	0.0209	3.9591	0.0196	0.0220	5.4517
5		0.0626	0.0781	4.4657	0.0644	0.0853	7.2965	0.0649	0.0968	11.5339
K = 3										
10		0.0369	0.0431	5.4888	0.0385	0.0465	8.4069	0.0384	0.0471	11.2734
15		0.0281	0.0331	7.3489	0.0277	0.0319	8.8557	0.0269	0.0305	10.5265
5		0.0721	0.0929	6.3544	0.0735	0.1029	10.5472	0.0694	0.0886	9.9542
K = 4										
10		0.0431	0.0530	8.4576	0.0426	0.0509	10.2590	0.0416	0.0494	12.1918
15		0.0273	0.0298	6.1387	0.0304	0.0352	11.2533	0.0289	0.0323	12.0653

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER N-137	2. GOVT ACCESSION NO. AD-A112 980	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) An Empirical Bayes Estimate of Multinomial Probabilities		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER TR #381
7. AUTHOR(s) Khursheed Alam Amitava Mitra		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0451
9. PERFORMING ORGANIZATION NAME AND ADDRESS Clemson University Dept. of Mathematical Sciences Clemson, South Carolina 29631		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 365-049
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Code 434 Arlington, Va. 22217		12. REPORT DATE February 1982
		13. NUMBER OF PAGES 12
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Multinomial distribution; Maximum likelihood; Empirical Bayes; Quadratic loss.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This paper deals with the problem of estimating simultaneously the parameters (cell probabilities) of $m \geq 2$ independent multinomial distributions, with respect to a quadratic loss function. An empirical Bayes estimator is given which is shown to have smaller risk than the maximum likelihood estimator for sufficiently large value of m , at all points of the parameter space outside its boundary.		

DD FORM 1473

EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

FILMED
4-8